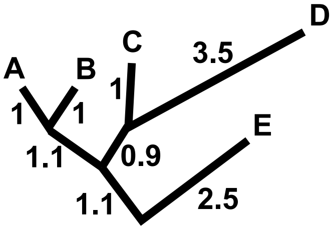
Exercise 1:



(a) Write this tree in Newick notation.

(b) Given this tree with additive branch lengths (representing evolutionary distances) calculate the pairwise distances of the five taxa.

(c) Carry out UPGMA and NeighborJoining with paper an pencil for this distance matrix.

(d) Carry out UPGMA and NeighborJoining with a computer program for this distance matrix.

Exercise 2: From the exact pairwise evolutionary distances of the taxa UPGMA can reconstruct each tree that fulfills the molecular-clock assumption from the distances of its leaves. Is this also true for hierarchical cluster methods that define the distance of two taxa groups as the minimum or the maximum pairwise distance instead of the mean distance that UPGMA is using? Give a proof or counterexamples.

Exercise 3: Install Phylip and R with the ape package on your computer and explore with both programs how well UPGMA and Neighbor Joining can estimate trees from imprecise distance data. Start with several distance matrices that are compatible to trees. How much can you perturb the distances until UPGMA and/or Neighbor Joining give you a different tree? Explore this with some trees that fulfill the molecular clock assumption and with some that do not.

Exercise 4: Work out the details of the proof of the neighbor joining theorem

(a) for the case that $2 \leq |L_1| \leq |L_2|$ (with the notations as in the lecture)

(b) for the case $1 = |L_1| \leq |L_2|$. For this case, show that if x is the only leaf in L_1 (that is, $L_1 = \{x\}$), neighbor joining would not join i and j as $D_{jx} < D_{ij}$.

Exercise 5: (For teams including at least one student with some experience in programming) Implement UPGMA or Neighbor Joining or both (in R, python, C, C++ or Java) and test your program with example data.

Exercise 6: Analyse how the time complexity of the neighbor-joining algorithm depends on the number of taxa (or on the size of the input data).

Exercise 7: Find a tree that explains the following sequence alignment without back-mutations or double-hits or explain why such a tree cannot exist. (Dots refer to the same nucleotide as in sequence A.)

```
Sequence A:  ACGTACGTTATCTTCATTGGTTCACATTCATGCGTATCAGTACG
Sequence B:  ...........G...................A............
Sequence C:  ..T..................C......................
Sequence D:  ..T..................C.........A..C.........
Sequence E:  ...............G.....C.........A.........G..
Sequence F:  .....................C.........A.........G..
```

Exercise 8: Find a perfectly parsimonious tree for the following sequence alignment or explain why such a tree does not exist. (Dots refer to the same nucleotide as in sequence U.)

```
Sequence U:  ACGTACGTTATCTTCATTGGTTCACATTCATGCGTATCAGTACG
Sequence V:  ....G.A.........C...........T..........C....
Sequence W:  ......A................C...............C....
Sequence X:  ................C...........T..........C....
Sequence Y:  .......................................C....
Sequence Z:  ...................................G...C....
```